

**PERBANDINGAN 3 METODE DALAM DATA MINING UNTUK  
PENENTUAN KADAR KOLESTEROL DI RSUD DR.MOEWARDI  
SURAKARTA**



**Disusun sebagai salah satu syarat menyelesaikan Program Studi Strata I pada Jurusan  
Informatika Fakultas Komunikasi dan Informatika**

**Oleh:**

**LIS MAHANANI**

**L 200 120 045**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**2016**

**HALAMAN PERSETUJUAN**

**PERBANDINGAN 3 METODE DALAM DATA MINING UNTUK  
PENENTUAN KADAR KOLESTEROL DI RSUD dr. MOEWARDI  
SURAKARTA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**PUBLIKASI ILMIAH**

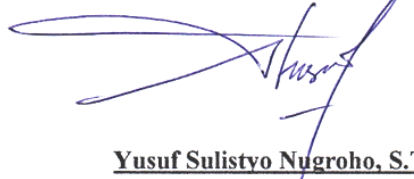
oleh:

**LIS MAHANANI**

**L 200 120 045**

Telah diperiksa dan disetujui untuk diuji oleh:

Dosen Pembimbing



**Yusuf Sulisty Nugroho, S.T., M.Eng.**

**NIK.1197**

**HALAMAN PENGESAHAN**

**PERBANDINGAN 3 METODE DALAM DATA MINING UNTUK  
PENENTUAN KADAR KOLESTEROL DI RSUD dr. MOEWARDI  
SURAKARTA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**OLEH**

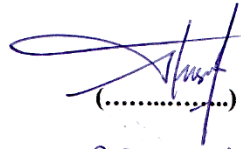
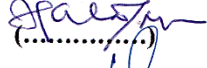
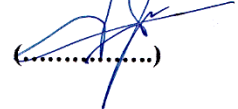
**LIS MAHANANI**

**L 200 120 045**

**Telah dipertahankan di depan Dewan Penguji  
Fakultas Komunikasi dan Informatika  
Universitas Muhammadiyah Surakarta  
Pada hari Rabu, 20 April 2016  
dan dinyatakan telah memenuhi syarat**

**Dewan Penguji:**

- 1. Yusuf Sulistyo Nugroho, S.T., M.Eng.  
(Ketua Dewan Penguji)**
- 2. Fatah Yasin Al Irsyadi, S.T., M.T.  
(Anggota I Dewan Penguji)**
- 3. Dedi Gunawan, S.T., M.Sc.  
(Anggota II Dewan Penguji)**

  
(.....)  
  
(.....)  
  
(.....)


**Publikasi ilmiah ini telah diterima sebagai salah satu persyaratan**

**Untuk memperoleh gelar sarjana**

**Tanggal 30-04-2016**

**Mengetahui,**

**Dekan  
Fakultas Komunikasi dan Informatika**  
  
**Husni Thamrin, S.T., M.T., Ph.D.**  
**NIK. 706**

**Kepala Program Studi  
Informatika**  
  
**Dr. Heru Suprivono, M.Sc.**  
**NIK. 970**

## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ada ketidakbenaran dalam pernyataan saya di atas, maka akan saya pertanggungjawabkan sepenuhnya.

Surakarta, 20 April 2016

Penulis



**LIS MAHANANI**

**L 200 120 045**



**UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA**

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id> Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

**SURAT KETERANGAN LULUS PLAGIASI**

**012/A.3-IL.3/INF-FKI/IV/2016**

Assalamu'alaikum Wr. Wb.

Biro Skripsi Program Studi Informatika menerangkan bahwa :

Nama : Lis Mahanani  
NIM : L200120045  
Judul : Perbandingan 3 Metode Dalam Data Mining Untuk Penentuan Kadar Kolesterol di RSUD Dr. Moewardi Surakarta.  
Program Studi : Informatika  
Status : Lulus

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Skripsi, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 27 April 2016

Biro Skripsi Informatika



Endang Wahyu Pamungkas, S.Kom., M.Kom.

Processed on: 27-Apr-2016 07:49 WIB  
ID: 665847159  
Word Count: 2796  
Submitted: 1

## PERBANDINGAN 3 METODE DATA MINING UNTUK PENEN...

By Lis-Mahanani

Similarity Index  
**30%**

## Similarity by Source

Internet Sources: 7%  
Publications: 2%  
Student Papers: 28%

PERBANDINGAN 3 METODE DATA MINING UNTUK PENENTUAN KADAR KOLESTEROL DI RSUD dr.MOEWARDI SURAKARTA  
Abstrak Rumah Sakit Umum Daerah dr.Moewardi menyediakan pelayanan untuk pemeriksaan penentuan kadar kolesterol yang diklasifikasikan menjadi tiga kategori yaitu kadar kolesterol normal, kurang dan lebih. Pemeriksaan ini bertujuan untuk mengetahui kadar kolesterol yang dimiliki oleh pasien. Kolesterol merupakan salah satu panel pemeriksaan di laboratorium RSUD dr.Moewardi Surakarta yang paling sering dilakukan. Sehingga dalam satu hari bisa mencapai seratus sampel pemeriksaan. Dengan data yang berlimpah dan berkelanjutan maka dapat dimanfaatkan sebagai

2 proses data mining. Teknik data mining yang digunakan dalam penentuan kadar kolesterol ini menggunakan metode Decision Tree Algorithm Index Gini, Naive Bayes dan

Rule Induction. Sedangkan atribut yang digunakan terdiri dari Jenis Kelamin, Usia, Riwayat Merokok, Indeks Masa Tubuh dan Kadar Kolesterol.

2 Dalam melakukan analisa ini menggunakan bantuan software RapidMiner 5 untuk mengetahui metode apa yang paling akurat. Pengimplementasian data mining menggunakan perbandingan 3 metode dapat diketahui bahwa berdasarkan dari nilai recall , metode

rule induction mempunyai nilai yang lebih tinggi dibandingkan

2 metode yang lain dengan nilai recall sebesar 95,12%. Berdasarkan nilai accuracy

1 11% match (student papers from 26-Apr-2016)  
Class Publikasi Wisuda Juni  
Assignment Publikasi Wisuda Juni  
Paper ID: [665479027](#)

2 8% match (student papers from 10-Mar-2015)  
Class publikasi  
Assignment publikasi  
Paper ID: [514648058](#)

3 4% match (student papers from 26-Apr-2016)  
Class Publikasi Wisuda Juni  
Assignment Publikasi Wisuda Juni  
Paper ID: [665479326](#)

4 1% match (Internet from 06-Feb-2015)  
<http://kadarkolesterolnormal.com>

5 1% match (student papers from 04-Feb-2014)  
Class publikasi maret 2014  
Assignment publikasi\_maret\_2014  
Paper ID: [393332890](#)

## PERBANDINGAN 3 METODE DATA MINING UNTUK PENENTUAN KADAR KOLESTEROL DI RSUD dr.MOEWARDI SURAKARTA

### Abstrak

Rumah Sakit Umum Daerah dr.Moewardi menyediakan pelayanan untuk pemeriksaan penentuan kadar kolesterol yang diklasifikasikan menjadi tiga kategori yaitu kadar kolesterol normal, kurang dan lebih. Pemeriksaan ini bertujuan untuk mengetahui kadar kolesterol yang dimiliki oleh pasien. Kolesterol merupakan salah satu panel pemeriksaan di laboratorium RSUD dr.Moewardi Surakarta yang paling sering dilakukan. Sehingga dalam satu hari bisa mencapai seratus sampel pemeriksaan. Dengan data yang berlimpah dan berkelanjutan maka dapat dimanfaatkan sebagai proses data mining. Teknik *data mining* yang digunakan dalam penentuan kadar kolesterol ini menggunakan metode *Decision Tree* Algoritma *Index Gini*, *Naive Bayes* dan *Rule Induction*. Sedangkan atribut yang digunakan terdiri dari Jenis Kelamin, Usia, Riwayat Merokok, Indeks Masa Tubuh dan Kadar Kolesterol. Dalam melakukan analisa ini menggunakan bantuan software RapidMiner 5 untuk mengetahui metode apa yang paling akurat. Pengimplementasian *data mining* menggunakan perbandingan 3 metode dapat diketahui bahwa berdasarkan dari nilai *recall*, metode *rule induction* mempunyai nilai yang lebih tinggi dibandingkan metode yang lain dengan nilai *recall* sebesar 95,12%. Berdasarkan nilai *accuracy* dan nilai *precision*, metode *naive bayes* mempunyai nilai lebih tinggi dibandingkan metode yang lain dengan nilai *accuracy* sebesar 54,83% dan *precision* sebesar 55,01%. Variabel yang paling berpengaruh dalam penelitian ini untuk menentukan kadar kolesterol pasien yaitu riwayat merokok.

**Kata kunci:** *Data Mining*, *Decision Tree* Algoritma *Index Gini*, Kolesterol, *Naive Bayes*, *Rule Induction*

### Abstract

dr.Moewardi public region hospital provides service for determining cholesterol levels checkup that classified into 3 categories they are , normal cholesterol, less and more. With cholesterol levels checkup, it can be known the cholesterol levels. Cholesterol is one of checkup panel in RSUD dr.Moewardi Surakarta. In one day it can reach one hundred of checkup samples. Continous data are plentiful and can be used for data mining process. Data mining techniques used in determining of cholesterol levels are *Decision Tree* Algoritma *Index Gini*, *Naive Bayes* and *Rule Induction*.method. While the attributes that are used such as gender, age, and relate a story of smoke, history of diabetes and cholesterol levels. RapidMiner 5 software is used to find out what the best method. The comparison of 3 methods can be calculated based on recall,Rule induction method has higher value than other method with recall value 95,12%. Based on accuracy value and precision, Naïve bayes method has higher value than other method with accuracy value 54,83% and precision value 55,01%. The most influential variable in determine cholesterol levels of patient is relate a story of smoke.

**Keywords:** Cholesterol, *Data Mining*, *Decision Tree* Algoritma *Index Gini*, *Naive Bayes*, *Rule Induction*

## 1. PENDAHULUAN

Kolesterol merupakan salah satu senyawa lemak yang lunak, berwarna kuning, yang diproduksi oleh tubuh, terutama di dalam hati. Setiap hari, hati menghasilkan sekitar 800 miligram kolesterol. Selain dari hati, kolesterol juga dapat berasal dari makanan produk hewani, makanan laut, susu, dan produk olahan susu. Kolesterol pada kadar normal berguna bagi kesehatan, kolesterol merupakan salah satu bentuk lemak penting yang diperlukan tubuh yaitu untuk pembentukan sel. Hembing (2008:1) dalam Purwaningtyas dan Kusumaningtyas (2011). Rumah Sakit Umum Daerah dr.Moewardi yang berdiri sejak tahun 1912 merupakan salah satu Rumah Sakit yang beralamat di Jalan Kolonel Sutarto 132, Surakarta. Di rumah sakit tersebut memberikan pelayanan untuk pemeriksaan penentuan kadar

kolesterol yang diklasifikasikan menjadi tiga kategori yaitu kadar kolesterol normal, kurang dan lebih. Kolesterol merupakan salah satu panel pemeriksaan di laboratorium rumah sakit umum daerah Dr.Moewardi Surakarta yang paling sering dilakukan. Sehingga dalam satu hari bisa mencapai seratus sampel pemeriksaan. Dengan demikian data sampel setiap harinya selalu meningkat. Menentukan kadar kolesterol yang dilakukan secara manual dinilai kurang efisien, karena data yang digunakan cukup banyak sehingga menyita waktu, membutuhkan ketelitian dan menguras tenaga.

Data mining merupakan proses untuk menganalisis data mentah dengan bantuan komputer dan mengekstraksi maknanya (Milovic, 2011). Dalam penelitian ini menerapkan proses *data mining*, selain untuk mendapatkan informasi penentuan kadar kolesterol diharapkan juga dapat menganalisis dari faktor – faktor yang paling berpengaruh dalam penentuan kadar kolesterol. Dengan demikian dapat dianalisis penentuan kadar kolesterol yang sudah ada serta menemukan informasi strategis dalam proses pengklasifikasian kadar kolesterol terhadap pasien. Berdasarkan permasalahan tersebut, maka dalam penelitian ini akan dilakukan perbandingan terhadap 3 metode untuk penentuan kadar kolesterol menggunakan *data mining*, yaitu metode *naive bayes*, *decision tree* algoritma *index gini* dan *rule induction*. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan *statistic* untuk memprediksi peluang di masa depan (Bustami, 2013). *Index gini* dipilih sebagai pengukuran *split* karena algoritma *entropy* memiliki operasi matematis yang lebih kompleks dibandingkan dengan metode berbasis *gini index* (Mandyartha, 2015). Menurut Kryszkiewicz dalam Dai (2011) metode *rule induction* digunakan untuk toleransi relasi dan mendefinisikan fungsi untuk sistem informasi yang tidak lengkap dan masing-masing tabel keputusan. Diharapkan setelah diterapkannya *data mining*, dapat memberikan informasi pemeriksaan penentuan kadar kolesterol dengan mudah bagi pasien serta memudahkan pihak rumah sakit untuk memberikan pelayanan pemeriksaan penentuan kadar kolesterol terhadap pasien secara efisien.

## 2. METODE PENELITIAN

### 2.1 Penentuan Variabel

Variabel yang akan digunakan dalam proses data mining ini ditentukan berdasarkan tujuan dari penelitian. Atribut-atribut yang digunakan yaitu seperti pada tabel.1.

**Tabel 1. Atribut yang digunakan**

Atribut	Variabel
Kadar Kolesterol	Y
Jenis Kelamin	X1
Usia	X2
Riwayat Merokok	X3
Indeks Masa Tubuh	X4

### 2.2 Pengumpulan data



Data-data yang digunakan adalah menggunakan data pasien pemeriksaan kolesterol tahun 2015 yang dimiliki oleh pihak RSUD dr.Moewardi Surakarta.

## 2.3 Penentuan sampel

Banyaknya data pasien yang diperoleh dari Rumah Sakit Umum Daerah dr.Moewardi Surakarta kemudian dihitung untuk menentukan jumlah sampel yang digunakan sebagai data pelatihan proses *data mining* dengan metode *Slovin* (Nugroho, 2014) seperti pada **persamaan 1**

$$n = \frac{N}{1+Ne^2} \dots \dots \dots (1)$$

Keterangan :

$n$  = jumlah sampel

$N$  = jumlah keseluruhan data / populasi

$e$  = galat kesalahan (sebesar 5%)

## 3.1 2.4 Naive Bayes

*Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan *statistic* untuk memprediksi peluang di masa depan. (Bustami, 2013)

Persamaan 2 teorema *Bayes* :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \dots \dots \dots (2)$$

## 3.2 2.5 Decision Tree Algoritma Index Gini

Mandiyartha, dkk (2015) menyatakan rumus index gini terdapat pada persamaan 2.

$$\text{Gini (A)} = 1 - \sum_{k=1}^C P_k^2 \dots \dots \dots (3)$$

## 3.3 2.6 Rule Induction

Hermawati (2013) menyatakan rumus induksi terdapat pada persamaan 4 dan 5.

a. Menghitung nilai *support*

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \dots \dots \dots (4)$$

b. Mencari nilai *confidence*

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \dots \dots \dots (5)$$

### 3. HASIL DAN PEMBAHASAN

#### 3.4 3.1 Pengambilan sampel

Apabila dalam satu tahun jumlah data pasien pemeriksaan kolesterol RSUD dr.Moewardi Surakarta memiliki jumlah 9000 pasien dan memiliki ketidakteelitian 5%. Maka jumlah sampel yang diambil yaitu :

$$n = \frac{N}{1 + Ne^2}$$

$$n = 9000 / 1 + 9000 \times (0,05)^2$$

$$n = 9000 / 1 + 9000 \times 0,0025$$

$$n = 9000 / 1 + 22,5$$

$$n = 9000 / 23,5$$

$$n = 382,9787234 \text{ pasien}$$

Jadi dibulatkan menjadi 383 pasien yang digunakan bahan *sampling*.

#### 3.5 3.2 Implementasi *Naive Bayes* menggunakan *RapidMiner 5*

Rancangan proses untuk prediksi data penentuan kadar kolesterol pasien menggunakan *RapidMiner 5* seperti pada gambar 1

Hasil *Naive Bayes* dengan tampilan *text view* seperti pada gambar 1.

```
SimpleDistribution
Distribution model for label attribute Kadar_Kolesterol

Class normal (0.535)
4 distributions

Class lebih (0.371)
4 distributions

Class kurang (0.094)
4 distributions
```

**Gambar 1. Tampilan hasil *Naive Bayes* pada *Text View***

Hasil *SimpleDistribution* seperti dalam gambar 1 menunjukkan model distribusi *Naive Bayes*. Berdasarkan hasil *naive bayes*, dapat dilihat bahwa model distribusi nilai kelas Normal sebesar 0.535, kelas Lebih sebesar 0.371 dan kelas Kurang sebesar 0.094.

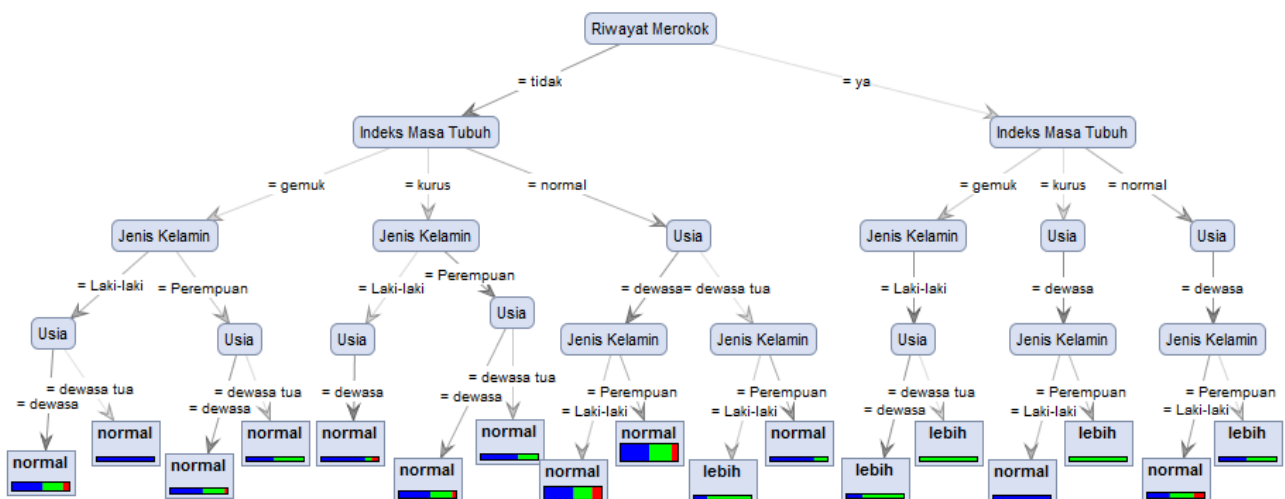
Hasil prediksi yang telah dieksekusi dapat dilihat pada gambar 2.

Row No.	confidence(...)	confidence(...)	confidence(...)	prediction(K...	Jenis Kela...	Usia	Riwayat Mer...	Indeks Mas...
1	0.530	0.371	0.099	normal	laki-laki	dewasa	tidak	gemuk
2	0.363	0.546	0.091	lebih	laki-laki	dewasa	ya	gemuk
3	0.347	0.499	0.155	lebih	laki-laki	dewasa	ya	normal
4	0.519	0.370	0.111	normal	perempuan	dewasa	tidak	normal
5	0.537	0.399	0.064	normal	perempuan	dewasa	tidak	gemuk
6	0.499	0.335	0.166	normal	laki-laki	dewasa	tidak	normal
7	0.519	0.370	0.111	normal	perempuan	dewasa	tidak	normal
8	0.519	0.370	0.111	normal	perempuan	dewasa	tidak	normal
9	0.355	0.543	0.102	lebih	perempuan	dewasa	ya	normal
10	0.519	0.370	0.111	normal	perempuan	dewasa	tidak	normal
11	0.618	0.305	0.077	normal	laki-laki	dewasa	tidak	kurus
12	0.624	0.327	0.050	normal	perempuan	dewasa	tidak	kurus
13	0.499	0.335	0.166	normal	laki-laki	dewasa	tidak	normal
14	0.363	0.546	0.091	lebih	laki-laki	dewasa	ya	gemuk
15	0.519	0.370	0.111	normal	perempuan	dewasa	tidak	normal
16	0.519	0.370	0.111	normal	perempuan	dewasa	tidak	normal
17	0.499	0.335	0.166	normal	laki-laki	dewasa	tidak	normal
18	0.519	0.370	0.111	normal	perempuan	dewasa	tidak	normal
19	0.530	0.371	0.099	normal	laki-laki	dewasa	tidak	gemuk
20	0.499	0.335	0.166	normal	laki-laki	dewasa	tidak	normal
21	0.519	0.370	0.111	normal	perempuan	dewasa	tidak	normal

Gambar 2. Hasil *Naive Bayes* untuk prediksi penentuan kadar kolesterol

### 3.6 3.3 Implementasi *Decision Tree* menggunakan *RapidMiner5*

Hasil skema pohon yang telah dieksekusi dari data penentuan kadar kolesterol menggunakan *RapidMiner 5* ditunjukkan pada gambar 3.



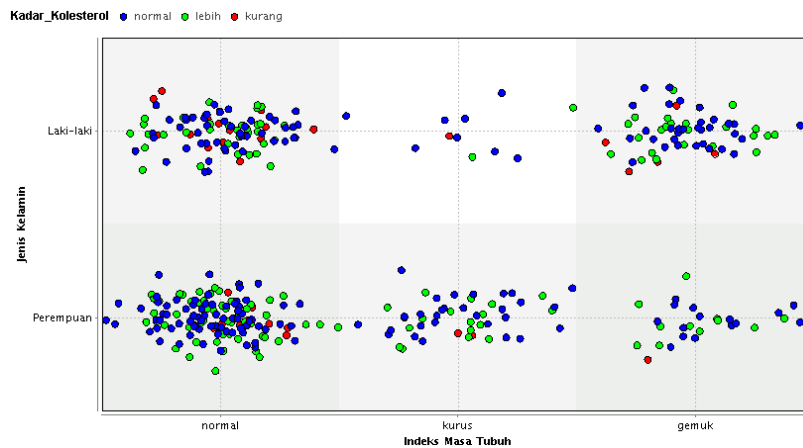
Gambar 3. Hasil skema *Decision Tree* untuk klasifikasi data pasien pemeriksaan kolesterol

Dari hasil pohon keputusan pada gambar 3 atribut Riwayat Merokok menempati posisi sebagai *root node*. Posisi internal node pertama pada Riwayat Merokok “tidak” dan Riwayat Merokok “ya” adalah atribut Indeks Masa Tubuh.. Semua variabel yang digunakan dalam penelitian ini mempengaruhi dalam penentuan kadar kolesterol pasien.

### 3.7 3.4 Implementasi *Rule Induction* menggunakan *RapidMiner 5*

Rancangan proses untuk memprediksi kadar kolesterol pasien menggunakan metode *Rule Induction* menggunakan *RapidMiner 5*. Hasil pada *scatter plot Rule Induction* berdasarkan data *training* dapat dilihat pada gambar 4.

Hasil pada *scatter plot Rule Induction* berdasarkan data *training* dapat dilihat pada gambar 4.



**Gambar 4. Scatter Plot Rule Induction**

Berdasarkan *plot view* pada gambar 4, penentuan kadar kolesterol pasien dengan indeks masa tubuh “gemuk”, “kurus” maupun “normal” dengan jenis kelamin “laki-laki” maupun “perempuan” menunjukkan bahwa kadar kolesterol pasien lebih dominan dengan kadar kolesterol “normal” dibandingkan dengan kadar kolesterol “kurang” dan “lebih”.

### 3.5 Implementasi Perhitungan *Naive Bayes*

Sebagai contoh diambilkan salah satu data uji yang memiliki ciri sebagai berikut : Laki-laki, usia dewasa, riwayat merokok tidak dan indeks masa tubuh gemuk. Apakah pasien tersebut mempunyai kadar kolesterol kurang, normal atau lebih?

Fakta menunjukkan :

$$P(Y = \text{Kurang}) = 0/50 = 0,067$$

$$P(Y = \text{Normal}) = 44/50 = 0,878$$

$$P(Y = \text{Lebih}) = 6/50 = 0,056$$

$$P(X_1 = \text{Laki-laki} | Y = \text{Kurang}) = 0/0 = 0$$

$$P(X_1 = \text{Laki-laki} | Y = \text{Normal}) = 16/44 = 0,364$$

$$P(X_1 = \text{Laki-laki} | Y = \text{Lebih}) = 5/6 = 0,833$$

$$P(X_2 = \text{Dewasa} | Y = \text{Kurang}) = 0/0 = 0$$

$$P(X_2 = \text{Dewasa} | Y = \text{Normal}) = 38/44 = 0,864$$

$$P(X_2 = \text{Dewasa} | Y = \text{Lebih}) = 5/6 = 0,833$$

$$P(X_3 = \text{Tidak} | Y = \text{Kurang}) = 0/0 = 0$$

$$P(X_3 = \text{Tidak} | Y = \text{Normal}) = 44/44 = 1$$

$$P(X_3 = \text{Tidak} | Y = \text{Lebih}) = 0/6 = 0$$

$$P(X_4 = \text{Gemuk} | Y = \text{Kurang}) = 0/0 = 0$$

$$P(X_4 = \text{Gemuk} | Y = \text{Normal}) = 8/44 = 0,182$$

$$P(X_4 = \text{Gemuk} | Y = \text{Lebih}) = 3/6 = 0,5$$

HMAP dari keadaan ini dapat dihitung dengan :

$P(X_1=\text{Laki-laki}, X_2=\text{Dewasa Tua}, X_3=\text{Ya}, X_4=\text{Ya}, | Y=\text{Kurang})$

$$= \frac{0}{0} \times \frac{0}{0} \times \frac{0}{0} \times \frac{0}{0} \times \frac{0}{50} = 0$$

$P(X_1=\text{Perempuan}, X_2=\text{Dewasa}, X_3=\text{Tidak}, X_4=\text{Tidak}, | Y=\text{Normal})$

$$= \frac{16}{44} \times \frac{38}{44} \times \frac{44}{44} \times \frac{8}{44} \times \frac{44}{50} = 0,001$$

$P(X_1=\text{Perempuan}, X_2=\text{Dewasa}, X_3=\text{Tidak}, X_4=\text{Tidak}, | Y=\text{Lebih})$

$$= \frac{5}{6} \times \frac{5}{6} \times \frac{0}{6} \times \frac{3}{6} \times \frac{6}{50} = 0$$

KEPUTUSAN PENENTUAN KADAR KOLESTEROL = **NORMAL**.

### 3.8 3.6 Implementasi Perhitungan *Decision Tree*

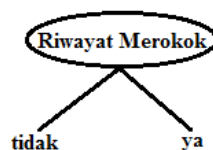
#### 3.6.1 Menentukan *Root Node*

Atribut yang memiliki nilai *gini split* paling rendah akan di pilih sebagai *root node*. Hasilnya dapat dilihat pada tabel.2.

**Tabel 2. *Gini Split* terendah**

Atribut	Nilai <i>Gini Split</i>
Jenis Kelamin	0,566
Usia	0,565
<b>Riwayat Merokok</b>	<b>0,562</b>
Indeks Masa Tubuh	0,564

Dari hasil seperti pada tabel 2 dapat disimpulkan bahwa atribut riwayat merokok menempati sebagai *root node*. Dapat dilihat pada gambar 5.



**Gambar 5. *Root Node***

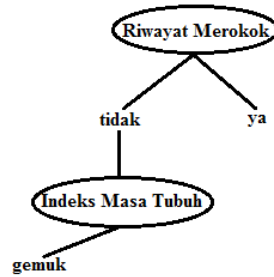
#### 3.6.2 Menentukan *Internal Node* pertama

Menentukan *internal node* pertama pada riwayat merokok “tidak” didapatkan nilai *gini split* seperti pada tabel 3.

**Tabel 3. Nilai *Gini Split* Indeks Masa Tubuh**

Nilai <i>Gini Split</i>	Riwayat Merokok “Tidak”
Jenis Kelamin	0,561
Usia	0,560
<b>Indeks Masa Tubuh</b>	<b>0,559</b>

Dari hasil seperti pada tabel 3 dapat disimpulkan bahwa atribut indeks masa tubuh merupakan *internal node* pertama pada riwayat merokok “tidak”, seperti pada gambar 6.



**Gambar 6. Internal Node Pertama**

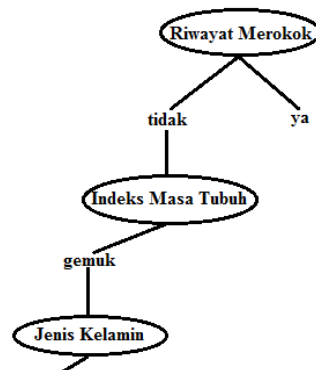
### 3.6.3 Menentukan *Internal node* kedua

Menentukan *internal node* kedua pada riwayat merokok “tidak” dan indeks masa tubuh “gemuk” hasilnya seperti pada tabel.4.

**Tabel 4. Nilai *Gini Split* riwayat merokok “tidak” dan indeks masa tubuh “gemuk”**

Nilai <i>Gini Split</i>	Riwayat Merokok “Tidak” dan Indeks Masa Tubuh “Gemuk”
Jenis Kelamin	0,542
Usia	0,543

Dari hasil pada tabel 4 dapat disimpulkan bahwa atribut jenis kelamin” merupakan *internal node* kedua pada riwayat merokok “tidak” dengan indeks masa tubuh “gemuk”, seperti pada gambar 7.



**Gambar 7. Internal Node Kedua**

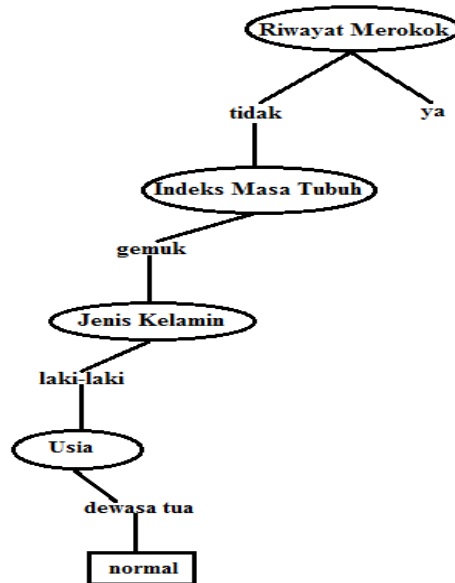
### 3.6.4 Menentukan *Internal node* ketiga dan *leaf node*

Setelah jenis kelamin sebagai *internal node* kedua maka usia sebagai *internal node* ketiga. Selanjutnya menentukan *leaf node* pada riwayat merokok “tidak”, indeks masa tubuh “gemuk”, jenis kelamin “laki-laki” dan usia “dewasa tua” hasilnya seperti pada table 5.

**Tabel 5. Nilai *Gini Split* Usia “Dewasa Tua”**

Nilai <i>Gini Split</i>	Riwayat Merokok “Tidak”, Indeks Masa Tubuh “Gemuk”, Jenis Kelamin “Laki-laki” dan Usia “Dewasa Tua”
Usia	0,000

Dari hasil tersebut dapat disimpulkan bahwa riwayat merokok “tidak, indeks masa tubuh”gemuk” dengan jenis kelamin “laki-laki” dan usia “dewasa tua” maka menghasilkan *leaf node* “normal” dikarenakan hasil dari semua *gini split* bernilai 0 dan probabilitas pasien dengan kadar kolesterol lebih dan kurang tidak ada.. Hasil pada gambar 8.



**Gambar 8. Leaf Node**

### 3.9 3.7 Implementasi Rule Induction

Berdasarkan pada *plot view* seperti pada gambar 4 bahwa kadar kolesterol “normal” lebih dominan terhadap indeks masa tubuh “gemuk”, “kurus” dan “normal” dengan jenis kelamin “perempuan”. Begitu juga dengan indeks masa tubuh “gemuk, “kurus” atau “normal” juga menghasilkan lebih banyak kadar kolesterol “normal”.

### 3.8 Hasil Perbandingan Metode

Untuk mengetahui tingkat keakuratan dari metode yang paling baik dalam penelitian ini yaitu *Decision Tree* algoritma *Index Gini*, *Naive Bayes* dan *Rule Induction*. Setelah semua metode dianalisa maka selanjutnya melakukan perbandingan dalam hitungan *Accuracy*, *Precision* dan *Recall*. Hasil perbandingan dapat dilihat pada tabel 6.

**Tabel 6. Hasil Perbandingan**

Komponen	<i>Decision Tree</i>	<i>Naive Bayes</i>	<i>Rule Induction</i>
<i>Accuracy</i>	53,26%	54,83%	54,31%
<i>Precision</i>	53,74%	55,01%	54,47%
<i>Recall</i>	94,63%	93,66%	95,12%

Berdasarkan hasil perbandingan pada tabel 6 dapat disimpulkan bahwa berdasarkan nilai *accuracy* dan *precision* metode *Naive Bayes* lebih baik digunakan karena *naive bayes* memiliki nilai *accuracy* dan *precision* lebih tinggi dari metode lainnya yaitu nilai *accuracy* sebesar 54,83% dan nilai *precision* sebesar 55,01%. Namun berdasarkan nilai *recall*, metode *rule induction* lebih baik digunakan karena dalam penelitian ini karena mempunyai nilai *recall* lebih tinggi dari

metode yang lainnya yaitu sebesar 95,12%. Metode *decision tree* mempunyai nilai *accuracy*, *precision* dan *recall* paling rendah dari metode *naïve bayes* dan metode *rule induction*.

### 3.10 3.9 Interpretasi Hasil Penelitian

Dari hasil analisis yang dilakukan, telah memberikan gambaran untuk menentukan kadar kolesterol dengan menggunakan berbagai metode. Metode yang digunakan dalam penelitian ini yaitu metode *Decision Tree* Algoritma *Gini Index*, *Naive Bayes* dan *Rule Induction*. Interpretasi hasil penelitian ini dijabarkan sebagai berikut:

#### 1. Interpretasi menggunakan metode *Decision Tree* Algoritma *Gini Index*

Dalam penelitian ini, Riwayat Merokok merupakan variabel yang paling mempengaruhi. Hal ini terbukti pada metode ini menyatakan bahwa riwayat merokok menempati posisi sebagai *root node*. Aturan yang terbentuk dari hasil penelitian bahwa :

- a. Jika riwayat merokok “tidak”, indeks masa tubuh “gemuk”, jenis kelamin “laki-laki” atau jenis kelamin “perempuan” dan usia “dewasa” atau usia “dewasa tua” akan menghasilkan kadar kolesterol “normal”.
- b. Jika riwayat merokok “tidak”, indeks masa tubuh “kurus”, jenis kelamin “laki-laki” atau jenis kelamin “perempuan” dengan usia “dewasa” atau usia “dewasa tua” juga sama menghasilkan kadar kolesterol “normal” seperti pada riwayat merokok “tidak” dengan indeks masa tubuh “gemuk”.
- c. Jika riwayat merokok “tidak”, indeks masa tubuh “normal” akan bergantung pada usia, pada usia “dewasa” akan bergantung pada jenis kelamin baik “laki-laki” maupun “perempuan” akan menghasilkan kadar kolesterol “normal”.
- d. Jika riwayat merokok “tidak”, indeks masa tubuh “normal” dan usia “dewasa tua” akan bergantung pada jenis kelamin, untuk jenis kelamin “laki-laki” akan menghasilkan kadar kolesterol “lebih” sedangkan jenis kelamin “perempuan” akan menghasilkan kadar kolesterol “normal”.
- e. Jika riwayat merokok “ya”, indeks masa tubuh “gemuk” dengan jenis kelamin “laki-laki” dan usia “dewasa” maupun “dewasa tua” maka akan menghasilkan kadar kolesterol “lebih”.
- f. Jika riwayat merokok “ya”, indeks masa tubuh “kurus” dengan usia “dewasa” dan jenis kelamin “laki-laki” akan menghasilkan kadar kolesterol “normal”, sedangkan dengan jenis kelamin “perempuan” akan menghasilkan kadar kolesterol “lebih”.
- g. Jika riwayat merokok “ya”, indeks masa tubuh “normal” dengan usia “dewasa” dan jenis kelamin “laki-laki” akan menghasilkan kadar kolesterol “normal” dan untuk jenis kelamin “perempuan” maka akan menghasilkan kadar kolesterol “lebih”.

#### 2. Interpretasi menggunakan metode *Naive Bayes*

Dalam penelitian ini menghasilkan prediksi yaitu

- a. Pasien dengan jenis kelamin “perempuan”, usia “dewasa”, riwayat merokok “tidak” dan indeks masa tubuh “normal” menghasilkan kadar kolesterol “normal”.
- b. Sedangkan pasien dengan jenis kelamin “laki-laki”, usia “dewasa”, riwayat merokok “ya” dan indeks masa tubuh “gemuk” menghasilkan kadar kolesterol “lebih”.



### 3. Interpretasi menggunakan metode *Rule Induction*

Algoritma *rule induction* melakukan perhitungan data set, namun pada data pasien pemeriksaan kadar kolesterol ini tidak ditemukan sequential item set. Hasil dari pengolahan data menunjukkan faktor yang mempengaruhi kadar kolesterol adalah riwayat merokok.

## 4. PENUTUP

Berdasarkan hasil penelitian maka dapat ditarik kesimpulan bahwa :

1. Riwayat Merokok merupakan variabel yang paling berpengaruh dalam penelitian ini. Hal ini terbukti pada metode *decision tree* yang menyatakan bahwa Riwayat Merokok menempati posisi sebagai *root node*.
2. Berdasarkan nilai *accuracy* dan *precision* metode *naïve bayes* lebih baik digunakan dalam penelitian ini dibandingkan dengan metode yang lain karena *naïve bayes* memiliki nilai *accuracy* sebesar 54,83% dan nilai *precision* sebesar 55,01%
3. Berdasarkan nilai *recall*, metode *rule induction* lebih baik digunakan dalam penelitian ini, karena mempunyai nilai *recall* lebih tinggi dari metode yang lain dengan nilai *recall* sebesar 95,12%.

## DAFTAR PUSTAKA

- Bustami. 2013. *Penerapan Algoritma Naive Bayes untuk Mengklasifikasi Data Nasabah Asuransi*, Jurnal Penelitian Teknik Informatika.
- Dai, J., Xu, Q., & Wang, W. 2011. *A Comparative Study on Strategies of Rule Induction for Incomplete Data Based on Rough Set Approach*. International Journal of Advancements in Computing Technology , Volume 3, Number 3.
- Hermawati, Fajar astuti. 2013. "Data Mining", andi. Yogyakarta.
- Mandiyartha, E.P., Kurniawan, M., & Perdana, R.S. 2015. *Identifikasi Sel Darah Merah Bertumpuk Menggunakan Pohon Keputusan Fuzzy Berbasis Gini Index*. Jurnal Buana Informatika, Vol 6, No 1.
- Milovic, B. (2011). *Usage of Data Mining in Making Business Decision*. YU Info 2012 & ICIST 2012, (pp. 153-157).
- Nugroho, Yusuf Sulisty. 2014. *Penerapan Algoritma C4.5 untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta*. Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST). Yogyakarta. November 2014. ISSN: 1997-911X.

Purwaningtyas, Essy: Kusumaningtyas, Rimma Hilda. 2011. *Penentuan Dosis dan Waktu Reaksi Optimum Ekstrak Buah Sawo (Achras zapota L.) Dalam Menurunkan Kadar Kolesterol*. Jurnal Penelitian Mahasiswi UNY, Vol VI, No.1.